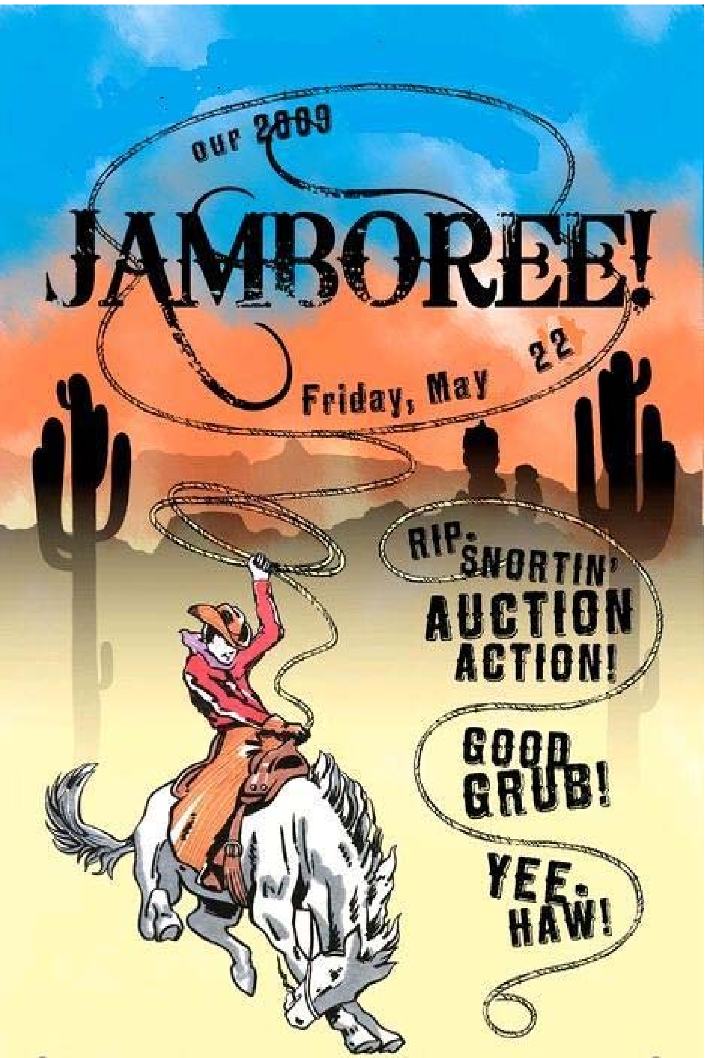


Run-, LumiBlock- wise “Conditions Data” and associated tools

Elizabeth Gallas - Oxford

ANL Analysis **“JAMBOREE”**

May 22, 2009

Preface:

I've never been to an "ATLAS jamboree".

My research into the term 'jamboree':

Wiktionary entry:

- ✓ **jamboree** (*plural* **jamborees**)
  - ➢ A lavish or boisterous celebration or party.

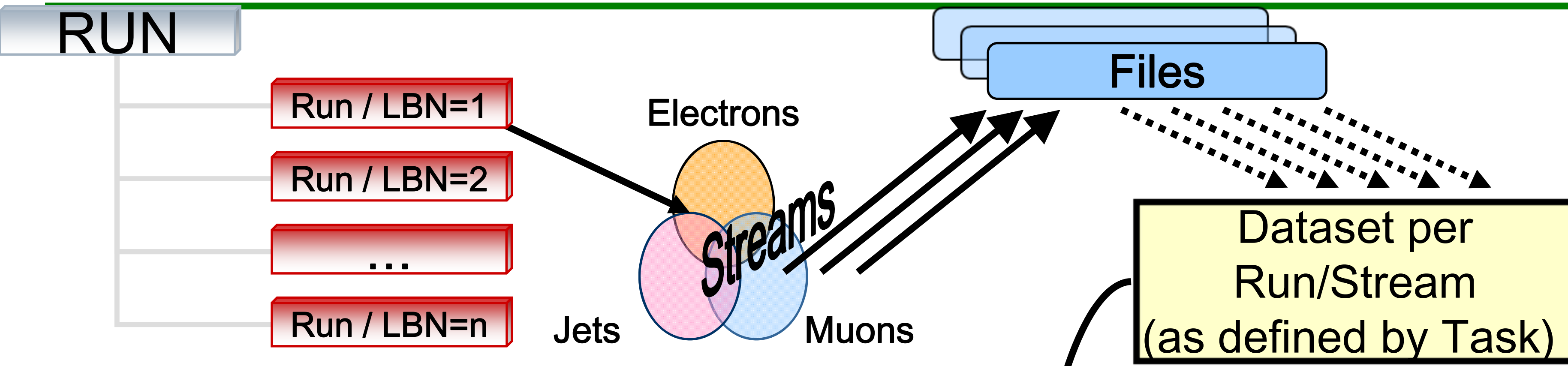
I image you all sitting in plastic lawn chairs with a cold drink (and your laptop).

All images displayed meet the google safe search criteria.

What's on the Menu ?

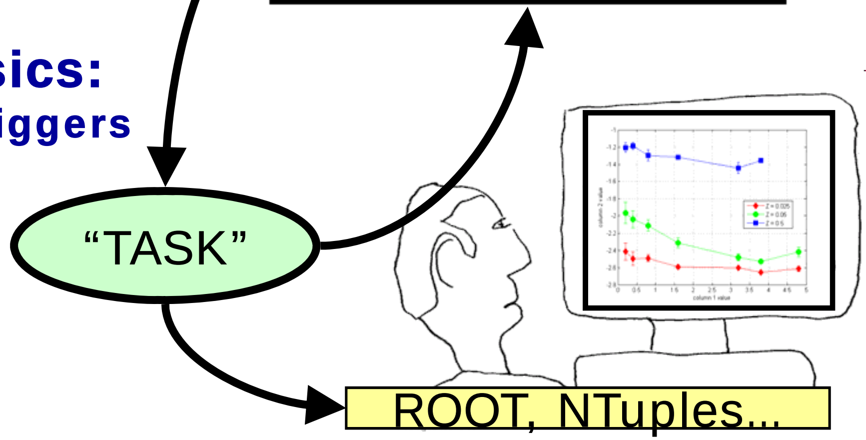
- Data path for physics events
  - Importance of complete samples
- Generalized "Task" description
- Data path for "Conditions"
  - The ATLAS Conditions Database
  - Stores a wide variety of conditions
- Luminosity
  - The Luminosity Working Group
  - Data sources and normalisation
  - Using tools to access conditions for studies and analysis
- Distribution of Conditions data on the grid
- Tools for getting samples of interest
  - Run-, Lumi-, and Event-wise criteria
- Summary
- Conclusion

Data Paths for (physics) “Events”



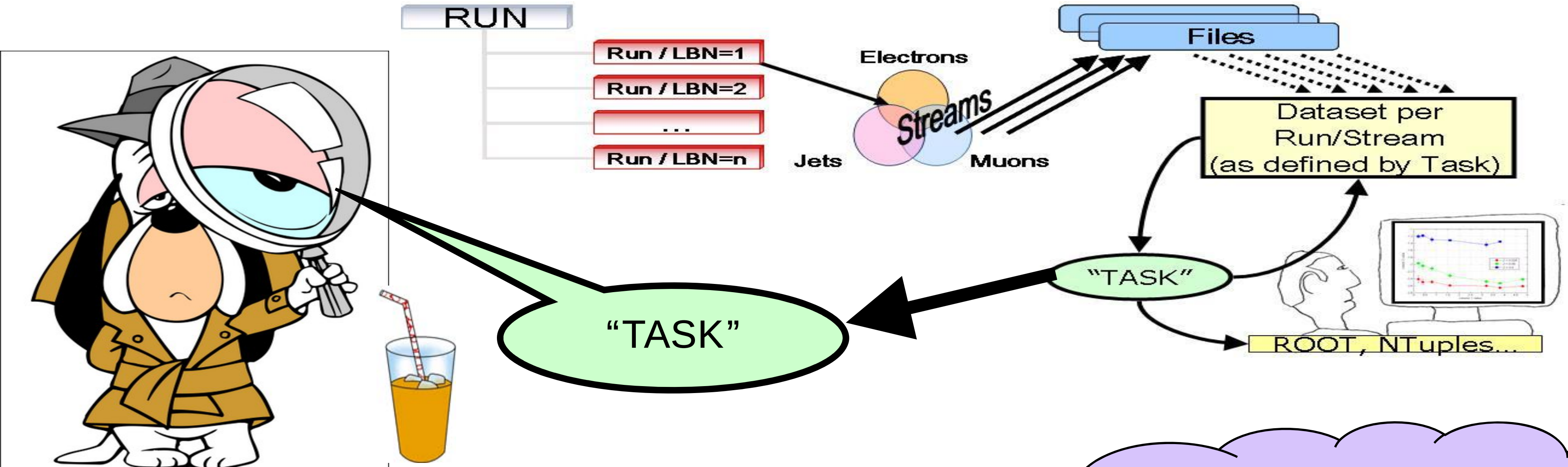
Ensuring complete datasets for physics:

- Include all events satisfying one / more triggers
- With dead time / losses precisely known
- In durations (in time) of known / constant
  - Luminosity and beam conditions
  - Detector / Trigger Configuration



- Events recorded in Runs (~hours)... divided into “Luminosity Blocks” (minutes)
- Events written to one/more Physics Streams ... based on trigger decision(s)
- Physics Streams are written to files ... respecting LB boundaries
- Files are processed into Datasets (careful accounting detect losses/failures)
- Track: Provenance (processing history); In-File metadata: list of analyzed LB’s.

Our many "Tasks"



Official Processing of Datasets is defined by a <u>TASK</u>

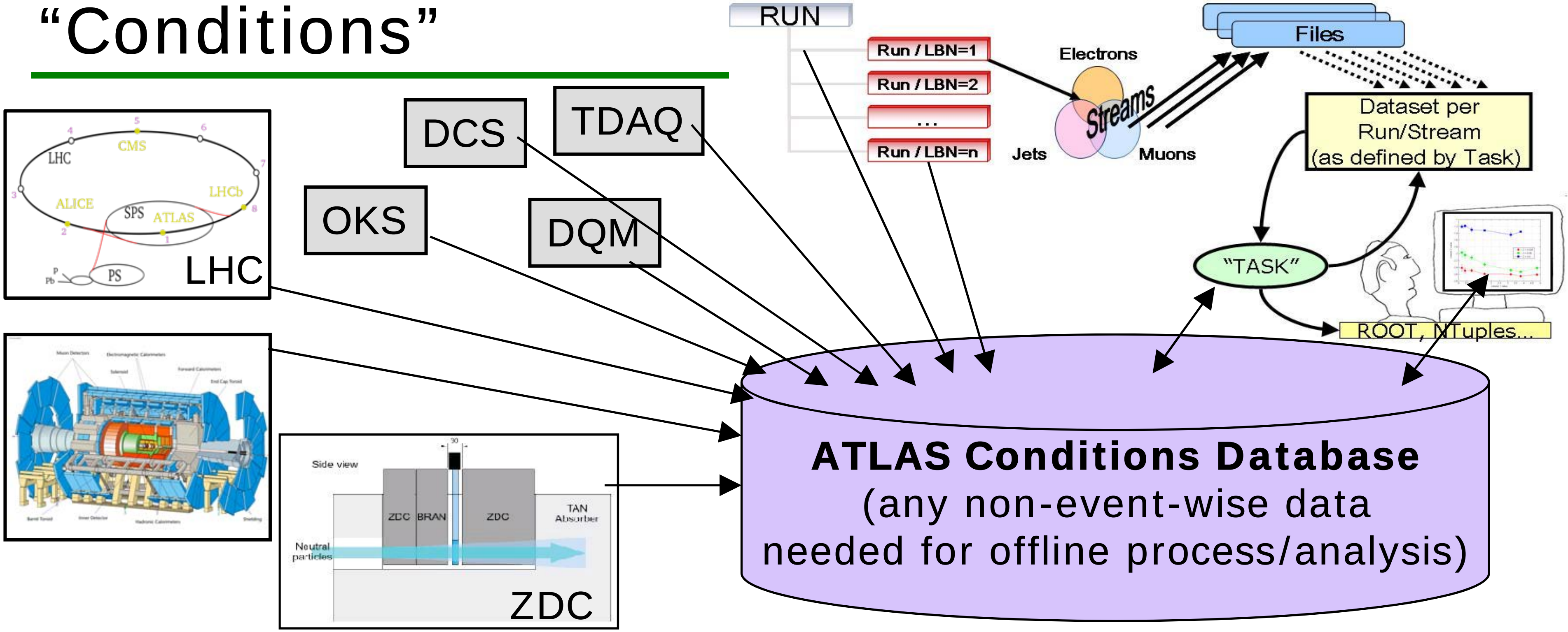
- Task definitions include Metadata pointers
  - To tagged versions of released software
  - And other **<u>tagged conditions</u>** information...
- Reconstruct fundamental quantities representing physics objects
- Produce new Datasets in various formats defined in the ATLAS computing model

Other processing of these and other data and datasets are used to

- Calibrate, align, ...optimize using known processes, simulation:
  - frequently from different sets of events or other information
- Produce new **<u>tagged conditions information</u>**

Where do these 'tagged conditions' fit in?

“Conditions”



“Conditions” – general term for information which is not ‘event-wise’ reflecting the conditions or states of a system – conditions are valid for an interval ranging from very short to infinity.

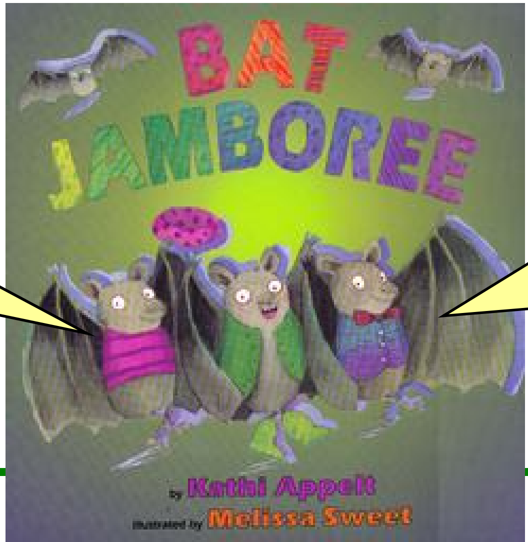
Any conditions data needed for offline processing and/or analysis must be stored in the

ATLAS Conditions Database

ATLAS Conditions Database

- Based on LCG Conditions Database infrastructure
  - Using 'COOL' (Conditions database Of Objects for LHC) API
  - Casually referred to as 'the COOL database' or 'storing in COOL…'
- COOL DB is an "Interval of Validity (IOV) Database"
  - Allowed interval - based on **Timestamps** or **Run/LB** ranges
- COOL 'Folders' -- Can be thought of as Tables
  - Indexed by: IOV, optional channel, optional COOL tag (a version)
  - Folder names are mnemonic and hierarchical (like UNIX pathnames)
    - ✓ indicating the relationship of one folder (or set of folders) to the next
- COOL Folder 'payload' == the data (the rows in the table)
  - Many storage options available within Conditions infrastructure
    - ✓ Optimized to the type of information being stored
  - Alternatively, the payload can be a pointer to an external structure (POOL file) or to another table.

Q: What conditions does the average bat need ?



A: That depends on what you are doing…

Conditions for the (a)typical bat:

- Common tasks: standard tools make COOL DB transparent to users
  - ✓ Some existing methods described later
- For specialized tasks:
  - COOL DB is readable in Athena: **CoolAthena** and **AthenaDBAccess**
  - You can make a ROOT file from COOL: **AtlCoolCopy**

- Where is documentation on COOL Folder content ?
  - Folders are generally described on TWiki pages
    - ✓ Good example: TWiki: **CoolOnlineData**
  - Poke subsystem experts

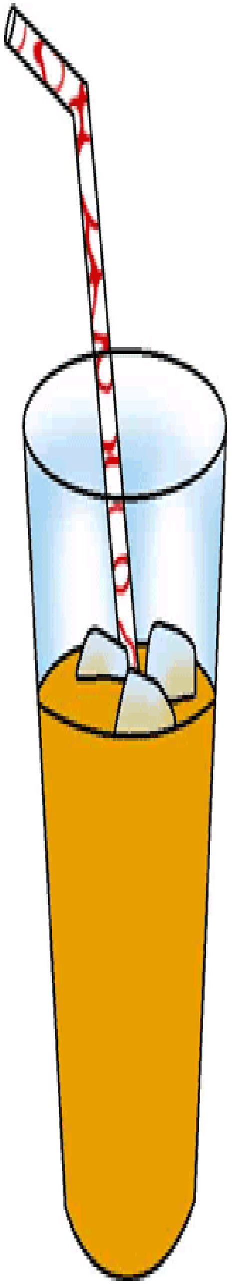
- A sampling of information stored in COOL:
  - LHC beam conditions
  - **online configuration and operation ***
  - calibration
  - alignment
  - **data quality ***
  - **luminosity and normalization ***
  - **object reconstruction efficiency *** and
  - bookkeeping data
    - ✓ cross checks of data completeness/integrity.

I'm glad you mention luminosity: Where can I find that for my data sample ?

Progress in Luminosity area

- Luminosity Working Group (April 2008) Members appointed to represent their ATLAS communities
  - Co-conveners: Benedetto Giacobbe and Marjorie Shapiro
  - Run Coordinator: Witold Kozanecki
  - TDAQ: Thilo Pauly
  - Conditions DB: Elizabeth Gallas
  - LUCID: Jim Pinfold
  - ALFA: Per Grafstrom
  - ZDC: Sebastian White
  - LHC Machine: Helmut Burkhardt
  - ATLAS-LHC Liaison: Sigi Wenig
  - SM Min Bias: Craig Butar
  - SM QCD: Uta Klein
- E-group: luminosityGroup
- Twiki: LuminosityGroup
- Aim: Provide a physicist doing offline data analysis with the information and software tools required to determine the instantaneous and integrated luminosity for any data sample with sufficient data quality.

A Tall order



- Determine Instantaneous Luminosity
  - Online measurement of Relative luminosity
  - Absolute Calibration
  - Point 1 Luminosity Panel
  - Offline luminosity determination
- Deadtime, Losses and Failures
- Assess Luminosity DQM
- Get all related data into COOL
- Use of Lumi Info in Studies/Analysis
  - Develop integrated plan with DQ

- Many more details: see talk in ATLAS WEEK
  - 20-Feb-2009: M.Shapiro/B.Giacobbe

Cross Sections and Luminosity

$$\sigma \cdot BR = \frac{N - N_{bk}}{\alpha \cdot \varepsilon \cdot \int l \cdot p \cdot L dt}$$

Meta-data associated with dt

Corrections: Live-time l, pre-scales p, losses, failures

Luminosity measurements, integrated + instantaneous (bunch-by-bunch)

Time granularity dt
LumiBlocks (LB) ~ 1 min

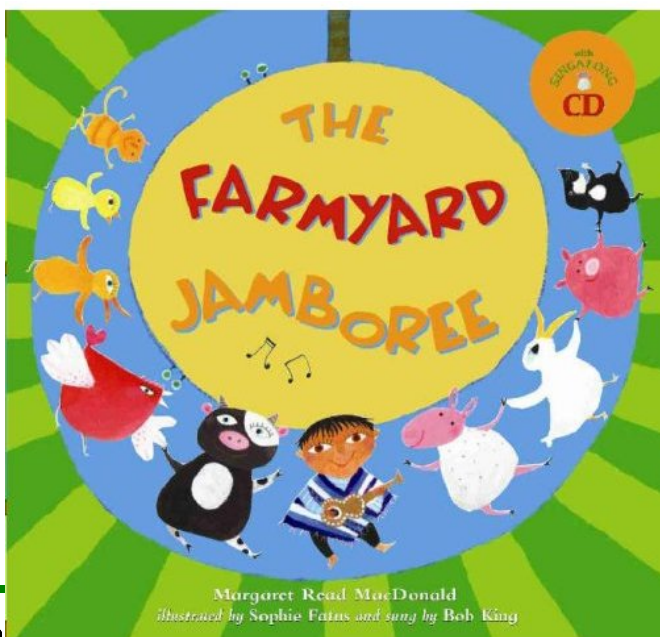
Luminosity WG provides oversight in all these areas

The Luminosity Farmyard

$$\int dt\, L(t) \cdot I(t) \cdot p(t) \cdot f(t)$$

Absolute Measurements

- Beam Parameters: 5 -10%
- ALFA : 2-3%
- Physics Processes with well calculable x-section
  - Pp-> ppee, ppμμ (low rate, eventually ~1-2% systematics?)
  - W/Z Production (high rate, but few % systematics)



Relative Measurements

- LUCID
- BCM
- ZDC
- MBTS
- L1 Calo Rates
- Min Bias Trigger Rate
- TileCal Anode, LAr Currents
- Offline physics signals
  - Pixel space points
  - Vertex counting
  - Resonance rates
- In all cases:
  - Monitor luminosity vs time (LB).
  - Calibrate to absolute measurements after the fact

Initial Strategy for Luminosity Determination

- Relative luminosity determined from ATLAS detectors:
  - LUCID
  - BCM
  - ZDC
  - MBTS
  - Others for monitoring purposes

$$\cdot \int dt\, L(t) \cdot I(t) \cdot p(t) \cdot f(t)$$

- Correlating results from different detectors and methods will allow us to assess systematics and sensitivity to background and acceptance
- Absolute calibration done via special runs w/LHC
  - Van der Meer Scan
- These calibrations will factor into the algorithm: provide real-time luminosity number based on measurements available
  - broadcast to LHC and detectors subsystems

Deadtime, Losses and Failures

$$\cdot \int dt\, L(t) \cdot I(t) \cdot p(t) \cdot f(t)$$

- Auditing Issues:
  - At each stage of processing must monitor number of input and output events to ensure no unexpected losses
  - Infrastructure in place for Level 1 Trigger
  - Tier 0 infrastructure in place to prevent processing if missing SFO files
  - HLT auditing strategy needs review
  - Analysis of failure modes for data reprocessing and user analysis in progress

Luminosity related measurements in COOL

- Work in progress: decisions on what to store factoring in what might be available over time and how that data might be used in the offline
  - For each subdetector, DCS, LHC
  - TDAQ (deadtime, prescales)
  - Wiki pages:
    - ✓ LuminosityOnlineSummary, LuminosityOnlineMonitor, LuminosityOnlineCool and LuminosityOffline
- Online:
  - Best online estimates: total and bunch-by-bunch
  - Store raw measurements for offline analysis
- Offline: luminosity calculated after the fact
  - Expect methods to improve as we learn more
  - Stored in COOL with COOL tag (version).
- Detector Status from DQM
  - Stored in COOL with COOL tag (version).

Additional help to meet the challenge …

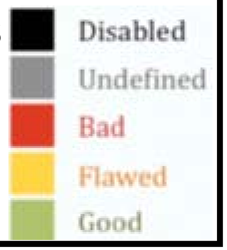
- Scope of task list – makes it interesting and fun
  - extends from LHC beam through the final analysis of events
    - ✓ And many things in between
  - Each task must be aware of integration issues up/down stream
  - Early analyses unaware of certain aspects may miss the mark
- Joining the band: Online, offline analysis thrill-seekers
  - Andrej Gorisek, David Berge, Stefan Maettig, Mika Huhtinen, Saverio D'Auria, Carla Sbarra, Antonio Sbizzi (… LUCID team), Slava Khomutnikov
  - Balint Radics, Soshi Tsuno, Akira Shibata, Regina Kwee, Kostas Nikolopoulos, Max Baak (and DQ coordinators)
- Advice/Help understand boundaries of infrastructure
  - Richard Hawkings, Giovanna Lehmann, David Malon
- Apologies to those I missed !!!

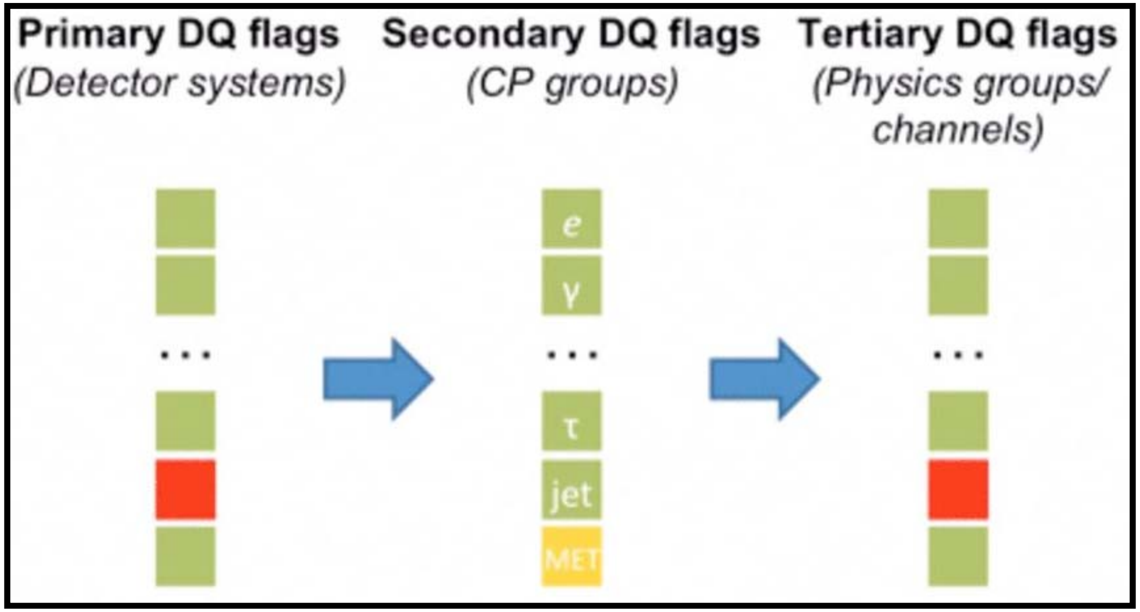
- Remains long list of tasks
  - Cross checks at every stage
  - Known unknowns
    - ✓ Each detector will have systematics
  - Unknown unknowns
    - ✓ Beam conditions…

DQ Group: Model for DQ assessments:

- Levels of assignment:
  - Primary
    - ✓ detectors and trigger slices,
  - secondary
    - ✓ CP groups
  - tertiary
    - ✓ physics channels - optional
- DQ Flags stored in COOL
  - With LB (or LB range) granularity
- Ongoing discussion on many issues

Allowed Flag Values





...What's the quality of your quality...?

Calculating Luminosity for your sample

Basic model:

- Stored in the data files: a list of analyzed LB's
  - ✓ Size of this record is modest for most analyses; Does not change with time
- Stored in the Database: Luminosity, Prescales, Deadtime, DataQuality
  - ✓ Too much to store in-file; Latest COOL tagged Lum, DQ... dynamically available

Modes of analysis:

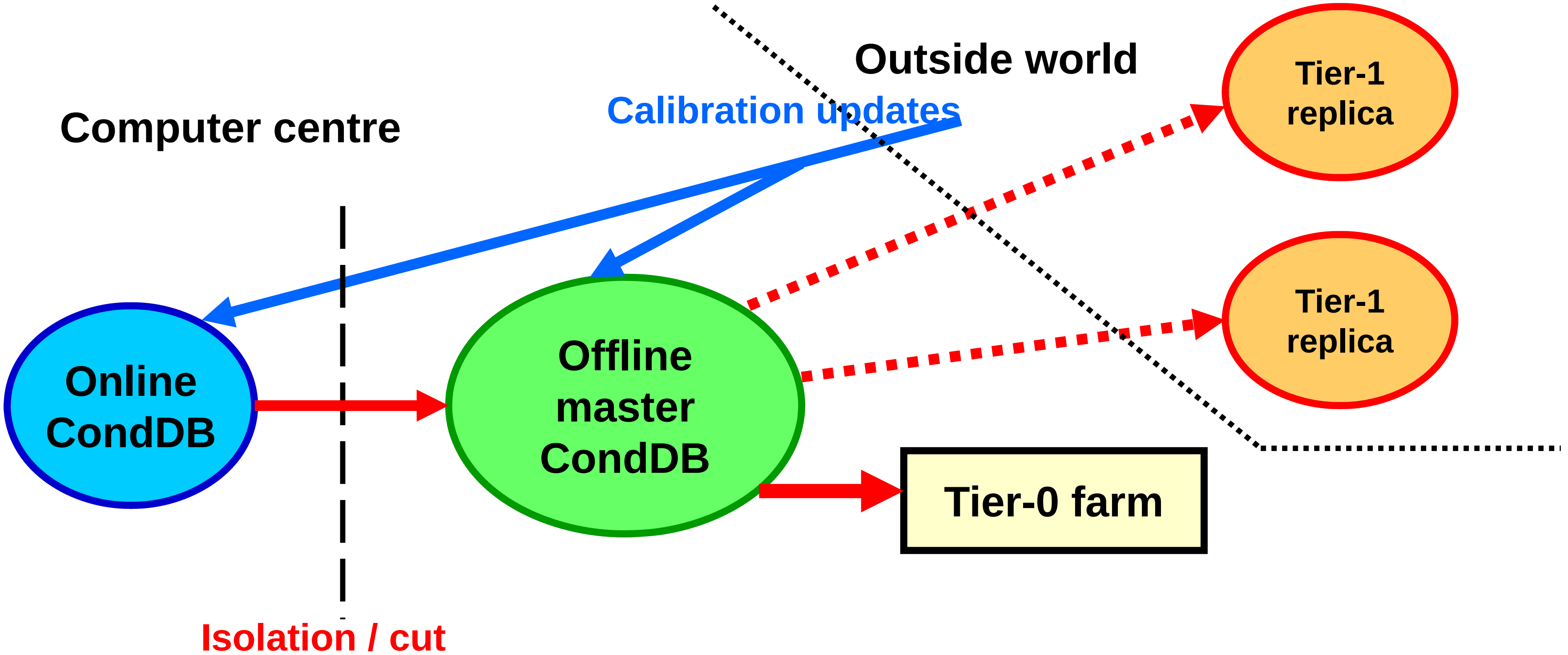
1. Each ESD/AOD/DPD file made in production or physics group SKIM contains in-file metadata: The list of LB's that have been processed
   - Same object stored in all 3 types of file
2. Query TagDB and iterate through EventCollection
   - Work underway to make this possible
3. Customized skimming:
   - make sure no events lost due to crashes (Missing events invalidates lumiCalc)
   - Must include this object record of all LB processed by your skimming job

Using LumiCalc.py: TWiki **CoolLumiCalc** (Prototyped: during FDR0,1,2)

- See Richard Hawkings talk 15-May-2008 in FDR Users meeting.
- Input: List of data files, Name of Trigger

• LumiCalc returns luminosity per file and total integrated luminosity
  - Corrects for deadtime and prescale factors
  - Extended to allow user to specify good run/LB list from DQ in COOL

Conditions Database replication: Tier-0 and Tier-1



- Master copy of ATLAS Conditions DB stored at CERN
- Data replicated to Tier 1
  - Using Oracle Streams technology
  - Can include non-COOL data (subdetector CORAL tables)
  - Only data needed for offline reconstruction/analysis
    - ✓ For example, some luminosity for studies only at Tier 0

Conditions DB distribution on the grid

Conditions Database access methods on the grid

- **Direct Oracle Access**
- **SQLite replica (files)** – create a ‘slice’ of Conditions data
  - ✓ Selected folders
  - ✓ Selected COOL tags
  - ✓ Selected IOVs
- **FroNTier/Squid (web) caching** technology (developed by CMS)
  - ✓ Being tested for use in ATLAS for various use cases

Each technology has advantages/disadvantages

Use cases dictate which technology is best

Trying to flesh out use cases for distributed analysis by users on the grid:

- Pre-knowledge of the input data needed by a job (production tasks)
  - Use Metadata to create local instance of data on/near where the job will be run
    - ✓ Sqlite files could contain: latest Luminosity, Data Quality and Efficiency…
- Ad-hoc queries (generally user tasks), where the task is less orchestrated:
  - Tier 1 RACs on multicore servers make ad-hoc queries more performant
  - Metadata, as well as the data it points to may be in the database or the Conditions DB reference is the Metadata is used to find the data
  - Metadata is used to retrieve the required data from the closest location

How do ATLAS physicists find Runs/Events of interest ?

Physicists have broad interests/responsibilities in ATLAS.
They need to find and analyze events offline.
How do they find Runs and Events for their purpose ?

Finding Runs of interest:

Examples:

- Sub-detector experts looking for cosmic ray data
  - ✓ And they'd like events with particular subdetectors engaged …
- Physicist in Group X wants to find events in Runs
  - ✓ designated by Group X to be 'good'

Solution: RunQuery tool: http://atlas-runquery.cern.ch/

- Web based system for querying the Conditions Database
- Allows the user to find the Runs of interest satisfying various Conditions (time, Detectors configured, Detector status flags…)
- Find datasets in AMI: **http://ami.in2p3.fr/** with a replica on cern.ch soon
  - ✓ Find your datasets, their provenance, their configuration tag meaning
  - ✓ Writes your dq2-get for you…

Finding Events of interest:

Example:

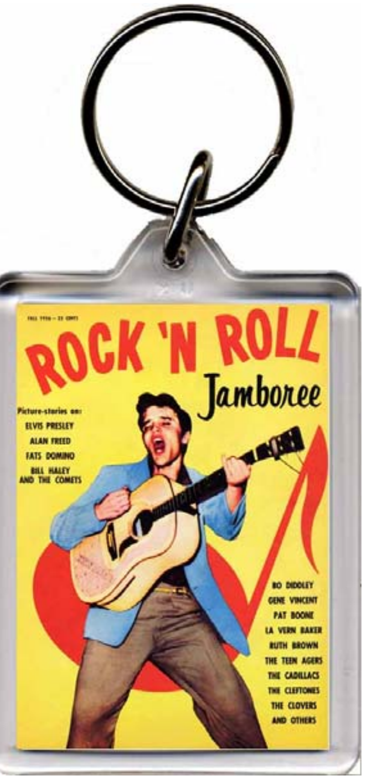
- Physicist wants to select events with offline electrons with pT > 30 GeV…

This is the basis for the ATLAS TAGs Application (next slide)

Metadata for Users: ATLAS TAGs

- PURPOSE: Facilitates event selection for analysis
- Available in File and Database formats (Storage: kB/event, >1TB/year)
  - Technical challenges in Poster on ATLAS TAGs distribution/management
- 'TAG Database' Application includes
  - Event-level Metadata produced routinely in data processing campaigns
    - ✓ About 200 indexed variables for each event: Identification keys, global event quantities, Trigger decisions, number of reconstructed objects (with their pT, eta, phi for highest-pT objects), Detector status,quality, physics, and performance words….
  - 'Run Metadata' at Temporal, Fill, Run, LB levels (from Conditions)
    - ✓ Has potential to add improved information (after ESD/AOD/TAG prod)
      - **Data Quality assessments**
      - **Efficiency calculations**
      - **Luminosity corrections**
  - References to Files for back-navigation
  - A variety of supporting tools and infrastructure
- Various components ATLAS TAG application are described
  - In this meeting
  - Recent CHEP presentations.
  - All software tutorials feature session on TAG usage

Summary



- Conditions DB infrastructure and distribution
  - Supports the conditions needed for processing, reprocessing, calibrations, alignments and vast array of offline studies and analysis
  - Distribution challenges addressed successfully so far
    - ✓ Use cases for analysis will expand with time
- Luminosity effort making headway on large task list
  - Much of the data to be stored as 'conditions' in COOL
    - ✓ Tools being developed accordingly
  - Significant work still to be done: more hands needed
  - Prototype Tool for Physics Users already exist
    - ✓ Exercised in Streaming Test and FDR
- Increasing array of tools making Conditions more accessible

Back Up Slides

Links

- Conditions Database users guide:
  - http://atlas.web.cern.ch/Atlas/GROUPS/DATABASE/project/online/doc/conddb_conventions.pdf
- LCG TWikis:
  - https://twiki.cern.ch/twiki/bin/view/LCG/COOL
  - https://twiki.cern.ch/twiki/bin/view/LCG/PyCool
- ATLAS TWikis:
  - ConditionsDB (Main)
  - CoolATLAS
  - AthenaDBAccess
  - CoolAthena
  - CoolPython
  - AtlCoolConsole
  - AtlCoolCopy
  - CoolUtilities
  - CoolProdAcc
  - CoolConnections
  - CoolTagging
  - CoolPublishing
  - CoolProdTags
  - CoolFileMetaData
  - CoolTools
  - InDetAlignHowTo
  - CoolDCS (PVSS-2-COOL)
  - Cool2Root ?
  - ...

Offline Determination of Luminosity

$$\cdot \int dt\, L(t) \cdot I(t) \cdot p(t) \cdot f(t)$$

- Determination of luminosity from LUCID non-trivial:
  - Must understand acceptance, backgrounds
  - At beginning will take time
  - Dedicated work required by LUCID group
  - Studies of alternative algorithms underway
    - ✓ see for example talk by M. Bruschi at LumiGroup meeting
- Compare LUCID determination with results from other detectors
  - evaluate systematics
- The luminosity group is responsible for providing “best” value of luminosity per LumiBlock:
  - Work just beginning in this area
  - In process of identifying point-person for each detector
  - Work must be coordinated with luminosity DQM
  - More manpower necessary

Luminosity Data Quality Monitoring Goals

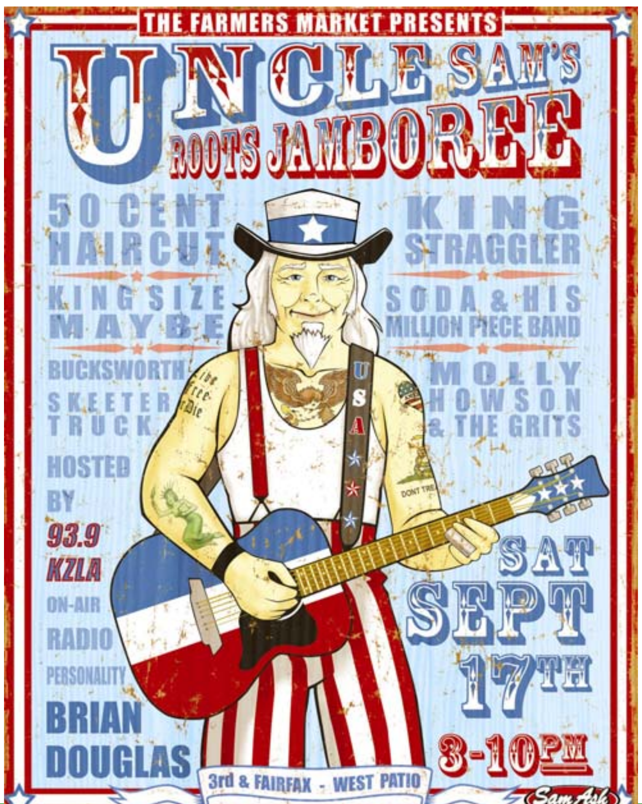
- Evaluate systematic uncertainties on lum measurements
  - Backgrounds
  - Nonlinearities
  - Changing beam conditions
- Flag any bad data that makes it through online checks
- Monitor wide range of rates that should scale with luminosity
  - Study correlations between different rates to evaluate systematics and separate detector problems from luminosity problems
- Two categories of information:
  - High rate triggers and currents that provide information per LumiBlock
  - Lower rate physics processes that provide information integrated over longer time scale (eg a full run)

Offline Data Quality Monitoring : Operational Strategy

- Fall 2009 effort will concentrate on offline monitoring via CAF
  - Need real data to understand where the real problems are
- As much as possible, piggy-back on existing efforts:
  - Min bias trigger and analysis
  - Inner Detector Monitoring
  - Forward detector analyses
  - At present, separate offline algorithms exist for monitor plot:
- Must integrate into single analysis job and develop operational model
- Goal for 2010 to migrate to Tier 0 DQM
- One complication: Must combine analyses from several stream and from COOL
  - Work required here

Metadata Bookkeeping for DPD's

- DPDs created from AODs for specific stream
- Some jobs may fail:
  - Core dumps
  - Some files not delivered
- Good News: Since AOD files are only closed on LB boundaries
  - We can calculate luminosity with partial dataset
  - But some LB may not have events in our AOD:
  - Need record of all LB processed by your skimming job
- Solution to bookkeeping problem:
  - Store the list of processed LB as in-file metadata
  - This works as long as job failures are properly handled


THE FARMERS MARKET PRESENTS
UNCLE SAM'S
ROOTS JAMBOREE
50 CENT HAIRCUT
KING STRAGGLER
KING SIZE MAYBE
SODA & HIS MILLION PIECE BAND
BUCKSWORTH
SKEETER TRUCK
MOLLY HOWSON & THE GRITS
HOSTED BY
93.9 KZLA
ON-AIR RADIO PERSONALITY
BRIAN DOUGLAS
SAT SEPT 17TH
3-10PM
3rd & FAIRFAX - WEST PATIO